Feature-Optimized Signal Quality Assessment in Wearable PPG Monitoring

Pietro Chierico[1], Daniel Suárez[2], Lorenzo Fácila[3],
Fernando Hornero[4], Raúl Alcaraz[5], José J Rieta[1]

[1] BioMIT.org, Electronic Engineering Department, Universitat Politecnica de Valencia, Spain
[2] Benejúzar Health Center, Orihuela Health Department, Spain
[3] Cardiology Department, General University Hospital Consortium of Valencia, Spain
[4] Cardiovascular Surgery Department, Hospital Clínico Universitario de Valencia, Spain
[5] Research Group in Electronic, Biomedical and Telecommunication Engineering, University of Castilla-La Mancha, Spain

Abstract

Continuous monitoring of photoplethysmographic (PPG) signals is crucial for assessing physiological parameters such as heart rate, pulse rate variability, and blood pressure. However, low-quality signal segments significantly impact measurement accuracy. In this study, we propose a novel signal quality assessment (SQA) algorithm based on boosted decision trees to classify PPG signal quality. The dataset, collected from 50 participants using a Polar Verity Sense device, consists of 7,402 manually labeled 10-second segments. A comprehensive feature extraction process yielded 48 attributes from time-domain, frequency-domain, and fiducial point-based analyses. After evaluating multiple machine learning models, the ensemble boosted tree classifier achieved the highest accuracy (98.12%). Feature selection using the Minimum Redundancy Maximum Relevance (MRMR) method reduced the feature set to 16 key attributes while preserving classification accuracy (97.0%). The results demonstrate that an optimized machine learning approach can provide robust and computationally efficient signal quality assessment.

1. Introduction

Continuous monitoring of photoplethysmographic (PPG) signals enables the calculation of various physiological parameters, such as heart rate (HR), Pulse Rate Variabilty (PRV), Blood Pressure (BP) and heart rate variability (HRV) [1]. However, the presence of low-quality signal segments can significantly degrade the accuracy of these measurements [2, 3]. Currently, no standardized algorithm exists for automatically assessing PPG signal quality and filtering out low-quality segments. To address this limitation, an automated approach is required to enhance the reliability of PPG-based physiological assessments.

Signal Quality Assessment (SQA) has been extensively studied in the scientific literature using a variety of methodologies, including signal processing techniques and machine learning-based approaches[4]. Recent research has introduced new flowchart-based frameworks for computing Signal Quality Indices (SQIs) [5, 6], many of which rely heavily on fiducial points extracted from the signal. Consequently, the accuracy of SQIs is strongly influenced by the choice of the beat detector. However, there is currently no universally accepted gold standard for beat detection [7], introducing variability and potential biases in signal quality estimation.

To overcome these limitations, alternative approaches have been explored, including the use of time-domain and frequency-domain features, as well as template-matching techniques [8, 9]. These methods allow for a more comprehensive evaluation of signal quality by capturing broader characteristics of the signal beyond fiducial point accuracy.

In this study, we propose a novel machine learning (ML) algorithm based on boosted decision trees for automated PPG signal quality classification. The algorithm extracts 16 distinct features from the signal and achieves high classification accuracy, demonstrating its robustness for SQA. Unlike traditional methods that rely predominantly on fiducial points, our approach employs an ensemble learning strategy that integrates features from all three domains: fiducial points, time domain, and frequency domain, ensuring a more reliable and generalizable signal quality assessment.

2. Materials and Methods

To develop and evaluate the SQA algorithm, a novel dataset of PPG signals was acquired. This dataset was collected using the Polar Verity Sense device (Polar Electro Oy, Kempele, Finland), an optical heart rate (OHR) sensor

 ISSN: 2325-887X DOI: 10.22489/CinC.2025.344

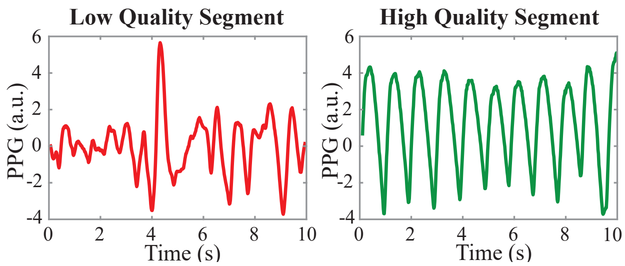

Figure 1. Examples of PPG signal segments with the two different quality levels: a low-quality segment on the left and an high quality segment on the right.

Table 1. List of the 48 extracted features to characterize the PPG categorized by the different domains.

Domain	Features
Time-Domain	Mean, Variance, Skewness, Kurtosis, Max, Min, Median, StdDev, Range, IQR, RMS PeakToPeak, ZeroCrossings, Perfusion, Entropy, Energy
Frequency-Domain	TotalPower, DominantFreq, Bandwidth, SpectralEntropy, LowFreqPower, HighFreqPower, RelLowFreqPower, RelHighFreqPower, SNR
PCPD Based Features	PeakCount, percentage, MeanPeak, VarPeak, MedianPeak, MaxPeak, MinPeak
Pulse-Pulse Interval (PPI)	MeanPPI, VarPPI, MinPPI, MaxPPI, HeartRate, SDNN, RMSSD, pNN50
MCC Based Features	MCC_Mean, MCC_Variance, MCC_Skewness, MCC_Kurtosis, MCC_Max, MCC_Min, MCC_Median, MCC_StdDev

operating at a sampling frequency of 55 Hz. In total, 20 hours recordings of PPG were obtained from 50 patients. The participant population had an average age of 53 years and included 60% women with 50% of participants engaging in regular physical activity, while the mean heart rate was 77 beats per minute (bpm).

A critical step in the development of the methodology was the accurate labeling of signal quality data. The raw PPG signals collected from participants underwent a preprocessing phase that included a Butterworth bandpass filtering with cutoff frequencies of 0.5 Hz and 8 Hz [10]. Following filtering, the dataset was divided into 10-second segments. This segmentation strategy was adopted to mitigate the challenge of assigning a single quality label to longer windows, such as 60 seconds, where signal quality can vary significantly within the same segment. By using shorter 10-second segments, fluctuations in signal quality could be more precisely captured, leading to an enhanced classification accuracy [11].

The final dataset comprised 20 hours of recordings, yielding a total of 7,402 segments. Each segment was manually labeled by three independent experts using a two-tier grading system. Segments were categorized as either high-quality or low-quality based on their visual clarity and morphological integrity. In the final distribution, 60% of the segments were labeled as high quality, while the remaining 40% were classified as low quality due to motion artifacts during signal acquisition. Figure 1 provides representative examples of PPG segments corresponding to each quality class.

The proposed algorithm is a machine learning classification model designed to assess signal quality using the described dataset, where the target variable represents the quality grade. This dataset comprised 48 features spanning multiple domains to comprehensively characterize the PPG signal, as detailed in Table 1.

To ensure a thorough signal characterization, five distinct categories of features are extracted. Time-domain features, such as skewness and median, capture the statistical distribution of the signal, while frequency-domain parameters, including total power and bandwidth, describe its spectral composition. A more refined representation is achieved through fiducial point-based features, derived using the Peakwise Correlation Pulse Detector (PCPD)[12], which is a robust algorithm specifically designed to identify valid peaks in the PPG. These features quantify the number of peaks, their median values, and additional metrics that characterize variations in the Pulse-Pulse Interval (PPI) across different recording segments [13].

To further enhance signal quality assessment, the algorithm incorporates the Minimum Correlation Curve (MCC), a fundamental component of PCPD. The MCC quantifies the correlation between the recorded PPG signal and high-quality reference windows, serving as a robust indicator of signal fidelity. By analyzing this curve, temporal descriptors, such as variance and kurtosis of the MCC are extracted, providing additional discriminative power to the classification model. This multi-domain feature extraction strategy ensures a robust representation of the PPG signal.

To ensure the highest quality of the dataset, standardization was performed using Z-score normalization, ensuring that all features have a mean of zero and a standard deviation of one, thereby enhancing the performance and stability of machine learning models.

Additionally, outlier detection was conducted using the IQR method, where values beyond the lower bound ($Q1 - 1.5 \times IQR$) or the upper bound ($Q3 + 1.5 \times IQR$) were identified and analyzed. The dataset was split into 70% for training and 30% for testing, while validation was performed using a ten-fold cross-validation approach to ensure model robustness and prevent overfitting. To tackle class imbalance in the target variable, the training dataset underwent undersampling, ensuring that the classes were perfectly balanced (50/50).

Using the structured and labeled dataset, multiple ML models were developed and evaluated. Each model was trained and validated using standard ML procedures to ensure robustness and generalizability. To further investi-

Table 2. Average classification accuracy for the different machine learning models tested in this study, listed in descending order of performance

Model Type	Accuracy (%)
Boosted Tree	98.12
Support Vector Machine	97.92
Neural Network	97.58
Efficient Linear SVM	97.56
K-Nearest Neighbors	97.21
Binary Logistic Regression	97.11
Linear Discriminant	97.08
Decision Tree	96.10
Naive Bayes	96.01

gate the discriminative power of different feature types, the dataset was also analyzed by feature domain: time-domain, frequency-domain, and fiducial point-based features, to determine whether specific domains were more effective in distinguishing between high- and low-quality PPG recordings.

Once the models were trained, the most suitable one, identified via the highest accuracy, underwent a feature reduction process using the Minimum Redundancy Maximum Relevance (MRMR) parameter. Analyzing this parameter with the elbow method made the reduction of features possible. This step aimed to remove insignificant features while identifying the relevant ones that significantly contribute to the target classification, thereby improving model interpretability and efficiency.

3. Results

Table 2 presents the validation accuracy of the ML models trained on the complete set of 48 features. The models exhibit high performance across both training-validation cross-folds and the testing phase, demonstrating their effectiveness in capturing relevant patterns in the data.

After evaluating multiple machine learning models, the Boosted Trees ensemble employing the GentleBoost algorithm was identified as the optimal approach. The base learners are decision trees with a maximum depth of 25 splits, and the ensemble consists of 50 learners. A learning rate of 0.05 was used to balance convergence speed and model stability.

The second analysis, conducted to assess the impact of different feature categories on the accuracy of the ensemble model, is shown in Figure 2. It is evident that peak-related features play a crucial role in model performance, achieving the highest accuracy among all categories. The left panel illustrates the classification accuracy of the best-performing model for each feature set, while the right panel presents the distribution of the 16 most relevant features selected via the MRMR parameter, highlighting the significance of these features in relation to the beat de-

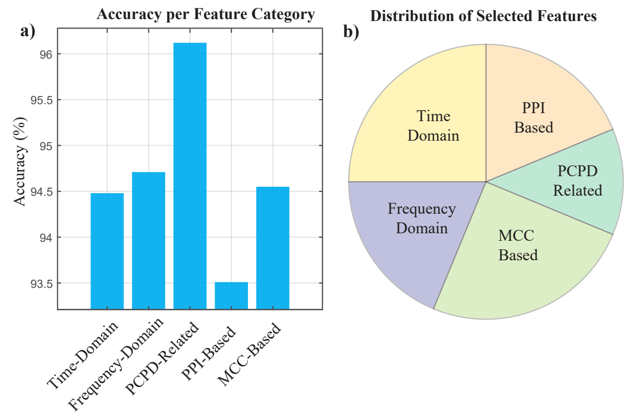

Figure 2. Relevance analysis of features. a) Classification accuracy for each feature category. b) Distribution of the 16 most relevant features selected by categories.

tection function. The final selected features, ranked by relevance using MRMR, included RMSSD, MCC Mean, VarPeak, MCC Min, Bandwidth, RelHighFreqPower, percentage, Kurtosis, Median, HighFreqPower, MinPPInterval, MCC StdDev, Skewness, SDNN, MCC Kurtosis, and Max, listed in descending order of importance.

The ensemble model, when trained with this refined set of 16 features, maintained a high level of accuracy, achieving 97.6% in the validation phase and 97.0% in the final testing phase. This demonstrates that a reduced yet highly informative feature set can sustain model performance while improving computational efficiency.

4. Discussion

The final accuracies achieved by the ensemble boosted tree model with the reduced feature set demonstrate that an optimal selection of features can sustain high classification performance while enhancing computational efficiency.

Among the most relevant features identified, RMSSD, MCC Mean, and MCC Min stand out. These metrics emphasize the significance of beat detection methodologies in the feature extraction process. In our case, the PCPD approach was employed, demonstrating its effectiveness in capturing essential signal characteristics without being influenced by the noise. The strong influence of these features on classification accuracy suggests that precise beat interval estimation plays a crucial role in distinguishing different physiological states.

Comparing our results with existing literature, we observe that different studies about SQA methods have reported similar classification performances. For instance, Liu et al. [14] reported an accuracy of 94% and 96% across training and testing phases, while Li et al. [8] attained 97% and 95%, respectiely. Additionally, Orphanidou et al. [5]

documented an accuracy of 97%, and Mohagheghian et al. [9] presented an average accuracy of 94%. These comparisons highlight that our model performs competitively within the state-of-the-art approaches.

A significant limitation in the field is the lack of a sufficiently large and standardized benchmark dataset with labeled PPG signal quality data. The widespread reliance on ECG as a reference standard introduces further complexity, as it requires synchronized multimodal recordings and may not reflect the standalone characteristics of the PPG signal. This dependency hinders the development of universally accepted datasets for evaluating PPG-based classification algorithms. Addressing this gap in future research would enable more consistent benchmarking and improve the generalizability of results across different studies.

5. Conclusions

This study presented a robust machine learning-based approach for PPG signal quality assessment, achieving high classification accuracy (97.0%) with an optimized feature set. The ensemble boosted tree model demonstrated superior performance, reducing dependency on fiducial points and improving computational efficiency. Future work should focus on developing standardized benchmark datasets to enhance comparability and generalizability across studies.

Acknowledgments

Research funded by grants PID2021-123804OB-I00, PID2021-00X128525-IV0 and TED2021-130935B-I00 from the Spanish Government and the European Regional Development Fund (10.13039/501100011033), SBPLY/21/180501/000186 from Junta de Castilla-La Mancha and AICO/2023/200 from Generalitat Valenciana.

References

[1] Charlton PH, Kyriaco PA, Mant J, Marozas V, Chowienczyk P, Alastruey J. Wearable photoplethysmography for cardiovascular monitoring. Proceedings of the IEEE Institute of Electrical and Electronics Engineers March 2022;110(3):355–381. ISSN 0018-9219.

[2] Scardulla F, D'Acquisto L, Colombarini R, Hu S, Pasta S, Bellavia D. A study on the effect of contact pressure during physical activity on photoplethysmographic heart rate measurements. Sensors January 2020;20(18):5052. ISSN 1424-8220.

[3] Allen J. Photoplethysmography and its application in clinical physiological measurement. Physiological Measurement March 2007;28(3):R1–39. ISSN 0967-3334.

[4] Vadrevu S, Manikandan MS. Real-time PPG signal quality assessment system for improving battery life and false alarms. IEEE Transactions on Circuits and Systems II Express Briefs November 2019;66(11):1910–1914. ISSN 1558-3791.

[5] Orphanidou C. Quality assessment for the photoplethysmogram (PPG). Signal Quality Assessment in Physiological Monitoring State of the Art and Practical Considerations 2018;41–63.

[6] Moscato S, Palmerini L, Palumbo P, Chiari L. Quality assessment and morphological analysis of photoplethysmography in daily life. Frontiers in Digital Health 2022; 4:912353.

[7] Charlton PH, Kotzen K, Mejía-Mejía E, Aston PJ, Budidha K, Mant J, Pettit C, Behar JA, Kyriacou PA. Detecting beats in the photoplethysmogram: benchmarking open-source algorithms. Physiological Measurement August 2022;43(8):085007. ISSN 0967-3334.

[8] Li Q, Clifford GD. Dynamic time warping and machine learning for signal quality assessment of pulsatile signals. Physiological Measurement August 2012;33(9):1491. ISSN 0967-3334.

[9] Mohagheghian F, Han D, Peitzsch A, Nishita N, Ding E, Dickson EL, DiMezza D, Otabil EM, Noorishirazi K, Scott J, et al. Optimized signal quality assessment for photoplethysmogram signals using feature selection. IEEE Transactions on Biomedical Engineering 2022;69(9):2982–2993.

[10] Chen J, Sun K, Sun Y, Li X. Signal quality assessment of PPG signals using STFT time-frequency spectra and deep learning approaches. In 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). ISSN 2694-0604, November 2021; 1153–1156. ISSN: 2694-0604.

[11] Orphanidou C, Bonnici T, Charlton P, Clifton D, Vallance D, Tarassenko L. Signal-quality indices for the electrocardiogram and photoplethysmogram: derivation and applications to wireless monitoring. IEEE Journal of Biomedical and Health Informatics May 2015;19(3):832–838. ISSN 2168-2208.

[12] Chierico P, Suarez D, Bertomeu-González V, Martínez-Rodrigo A, Alcaraz R, Rieta JJ. Peakwise Correlation Pulse Detector: A novel method for noise-resilient peak detection in PPG signals. In press. 2025; .

[13] Reddy GNK, Manikandan MS, Murty NVLN, Cenkeramaddi LR. Unified quality-aware compression and pulse-respiration rates estimation framework for reducing energy consumption and false alarms of wearable PPG monitoring devices. IEEE Access 2023;11:41708–41740. ISSN 2169-3536.

[14] Liu SH, Li RX, Wang JJ, Chen W, Su CH. Classification of photoplethysmographic signal quality with deep convolution neural networks for accurate measurement of cardiac stroke volume. Applied Sciences January 2020; 10(13):4612. ISSN 2076-3417.

Address for correspondence:

José J. Rieta
BioMIT.org, Electronic Engineering Department, Building 7F, Universitat Politecnica de Valencia, 46022 Valencia, Spain.
e-mail: jjrieta@upv.es